International Journal of Data and Network Science 10 (2026) 597–608

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Parameter estimation in mixed estimator nonparametric regression-spline truncated and fourier series (MENR-SF) for behavioral factors of prevalence of heart disease

I Nyoman Budiantara[a*], Nur Chamidah[b,c], Andrea Tri Rian Dani[d,e] and Muhammad Anshari[f]

[a]Department of Statistics, Faculty of Sciences and Data Analytics, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia
[b]Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Surabaya, Indonesia
[c]Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia
[d]Doctoral Study Program MIPA, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia
[e]Statistics Study Program, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia
[f]Business Information Systems at the Universiti Brunei Darussalam School of Business and Economics, Brunei Darussalam

C H R O N I C L E

Received October 8, 2025
Received in revised format December 25, 2025
Accepted January 29 2026
Available online January 29 2026

Keywords:
Mixed Estimators
Nonparametric Regression
Spline Truncated
Fourier Series
Prevalence of Heart Disease

A B S T R A C T

This study aims to develop and apply a Mixed Estimator Nonparametric Regression–Spline Truncated and Fourier Series (MENR–SF) to model the nonlinear relationships between behavioral factors and the prevalence of heart disease in Indonesia. The proposed approach simultaneously combines spline truncated estimators and Fourier series within a unified nonparametric regression framework, allowing each predictor variable to be modeled according to the specific characteristics of its relationship with the response variable. Parameter estimation is conducted using the Least Squares method, while the optimal number of spline knots and Fourier oscillations is determined based on the Generalized Cross-Validation (GCV) criterion. The application of the MENR–SF model to data from the 2023 Indonesian Health Survey (Survei Kesehatan Indonesia, SKI), with 38 provinces as the units of analysis, indicates that the best-performing model is obtained when the prevalence of daily smoking, the proportion of insufficient physical activity, and habitual consumption of fatty foods are modeled using spline truncateds, whereas the proportion of hypertension control is modeled using a Fourier series. The optimal combination, with three spline knots and three Fourier oscillations, yields a minimum GCV value of 1.197, low prediction error, and a coefficient of determination of 0.94, indicating an excellent ability of the model to explain variations in heart disease prevalence. These findings conclude that MENR–SF is a flexible and accurate approach for modeling complex nonlinear relationships in health data. The model offers enhanced flexibility and richer interpretability regarding the effects of behavioral factors, thereby holding strong potential to support data-driven health analysis and policy formulation.

1. Introduction

Nonparametric regression is a widely used statistical approach when the relationship between predictor variables and a response variable cannot be assumed to follow a specific functional form (Eubank, 1999). In contrast to parametric regression, which requires a predefined model structure, nonparametric regression offers a major advantage in its high flexibility to capture complex data patterns, including nonlinear relationships, local structural changes, and oscillatory behavior (Huang et al., 2013; Yatchew, 1998). This flexibility makes nonparametric regression highly relevant for modeling contemporary real-world phenomena (Zhang et al., 2023). Within nonparametric regression, several estimators are commonly employed, including spline (Chen, 1991; Eilers & Marx, 2010), kernel (Chu & Marron, 1991; Cui & Wei, 2013), and Fourier series estimators (Bilodeau, 1992; Ganesh et al., 2011), each with distinct characteristics. Spline estimators, particularly spline truncateds, are effective in modeling locally varying patterns across specific subintervals through the introduction of knot points (Budiantara, 2019). They can adaptively capture changes in trends; however, their performance is highly sensitive to

* Corresponding author
E-mail address: nyomanbudiantara65@gmail.com (I N. Budiantara)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2026.1.011

the number and placement of knots (Lestari et al., 2020). Kernel estimators, on the other hand, perform local smoothing and depend strongly on bandwidth selection, making them suitable for smooth relationships but less optimal for data exhibiting sharp changes (Hardle & Kelly, 1987). Fourier series estimators are effective in representing periodic or oscillatory patterns, yet they are less flexible in capturing nonperiodic local structural variations (Bilodeau, 1992; Greblicki & Pawlak, 1985).

As data complexity and the phenomena being modeled continue to increase, relying on a single type of nonparametric estimator for all predictor variables has become increasingly restrictive. The relationship between each predictor and the response variable often exhibits distinct characteristics, thereby requiring a more adaptive modeling strategy. This limitation has motivated the development of mixed-estimator-based nonparametric regression, which allows the simultaneous use of multiple estimators within a single nonparametric regression model. The development of mixed estimators in nonparametric regression was first introduced by (Budiantara et al., 2015) through a spline–kernel mixed estimator for a single predictor variable and was subsequently extended to the multivariate case by (Ratnasari et al., 2016, 2021). These approaches have been shown to enhance model flexibility and estimation accuracy by adapting to the characteristics of individual predictors. Nevertheless, further development of mixed estimators remains open, particularly in combining estimators capable of simultaneously capturing local patterns and oscillatory behavior.

Motivated by these considerations, this study proposes and develops a Mixed Estimator Nonparametric Regression–Spline Fourier (MENR-SF), an innovative nonparametric regression framework that integrates spline truncated and Fourier series estimators within a single modeling structure. The spline truncated estimator is employed to model nonlinear predictor–response relationships with local structural changes (Dani et al., 2021), while the Fourier series estimator is used to capture global and periodic oscillatory patterns (Amri et al., 2024). Model parameters are estimated using the Least Squares method, and the optimal selection of spline knot points and the number of Fourier oscillations is determined based on the Generalized Cross-Validation (GCV) criterion (Wahba, 1990).

A key novelty and main strength of this study lies in the fact that the estimator combination is not applied uniformly across all predictor variables but is instead selectively evaluated for each predictor. This strategy assumes that each predictor may exhibit a distinct relationship with the response variable in terms of smoothness, local structural variation, and fluctuating tendencies. Consequently, imposing a single estimator on all predictors may overlook their specific characteristics. Within the MENR-SF framework, each predictor variable is assessed to be modeled using a spline truncated estimator, a Fourier series estimator, (Wahba & Wang, 2014)or a combination of both, according to the empirical relationship patterns observed in the data. As a result, the constructed model is adaptive to predictor-specific characteristics, enabling it to capture more complex and realistic relationship structures. This approach not only enhances model flexibility but also has the potential to produce more accurate and interpretable estimates. The selection of the optimal estimator combination at the predictor level is conducted systematically using the GCV criterion, ensuring that the final model represents a global optimization of prediction error. This strategy extends previous mixed-estimator nonparametric regression approaches, which typically apply fixed estimator combinations, toward a more dynamic and data-driven framework.

The MENR-SF model is applied to data on prevalence of heart disease in Indonesia within the context of public health. Coronary heart disease is one of the leading non-communicable diseases (NCDs) exhibiting an increasing trend over time and imposing a substantial health burden in both developed and developing countries (Chamidah et al., 2025). According to the World Health Organization (WHO), approximately 41 million deaths worldwide in 2021 were attributed to NCDs, with heart disease accounting for the largest proportion about 43.6% or 17.9 million deaths. More than four out of five of these deaths were caused by heart attacks and strokes, and one-third occurred among individuals under 70 years of age, highlighting a serious global health threat (Kim & Johnston, 2011; Sanchis-Gomar et al., 2016). This global situation is consistent with conditions in Indonesia. Based on the Decree of the Minister of Health of the Republic of Indonesia No. HK.01.07/MENKES/1419/2023 and the 2024 Indonesian Health Profile (Kementerian Kesehatan RI., 2024), heart disease remains the leading cause of mortality and represents one of the highest-cost disease categories within the national health insurance system (BPJS Kesehatan), with approximately 22,550,047 cases annually (Kementerian Kesehatan RI., 2025). The high prevalence and associated healthcare expenditure underscore the urgent need for targeted attention, particularly in prevention, early detection, and data-driven risk factor control. Various behavioral and lifestyle factors such as unhealthy dietary patterns (Hajduk & Chaudhry, 2016), insufficient physical activity (Liu et al., 2025), and smoking habits (Hackshaw et al., 2018) play a major role in increasing the prevalence of heart disease. The complexity of these factors leads to predictor–response relationships that are inherently nonlinear, nonmonotonic, and potentially fluctuate. Therefore, a flexible and adaptive statistical modeling approach is required.

In this context, MENR-SF is expected to provide a more accurate and interpretable representation of predictor–response relationships compared with single nonparametric estimators. In addition to contributing to the methodological advancement of nonparametric regression, this study supports data-driven health policy development in line with the Sustainable Development Goals (SDGs), particularly Goal 3: Good Health and Well-Being, through more comprehensive modeling and analysis of health data.

2. Materials and methods

In the Materials and Methods section, single nonparametric estimators namely the spline truncated estimator and the Fourier series estimator are first introduced, as they constitute the main components in the construction of the proposed mixed estimator model. This section presents the mathematical formulations of each estimator along with their respective characteristics in capturing nonlinear relationship patterns. Furthermore, it describes the data sources, and the research variables employed in modeling prevalence of heart disease.

2.1. Nonparametric Regression Spline Truncated

A spline truncated is a segmented polynomial model developed to overcome the limitations of conventional polynomial functions, particularly their global nature (Fatmawati et al., 2019). Owing to its characteristics, the spline truncated provides a more flexible alternative that can accommodate changes in relationship patterns within the data more effectively than standard polynomial functions. The spline truncated estimator can address data patterns that exhibit sharp increases or decreases through the introduction of knot points, thereby enabling an accurate estimation of the regression curve (Ramli et al., 2020). Suppose that paired data $(z_{1i}, z_{2i}, \dots, z_{si}, y_i)$ are observed, where the relationship between the predictor variables and the response variable follows a spline truncated nonparametric regression model. The model can be expressed in Eq. (1).

$$y_i = h(z_{1i}, z_{2i}, \dots, z_{si}) + \varepsilon_i \tag{1}$$

where ε_i denotes the random error term. If the predictor variables $z_{1i}, z_{2i}, \dots, z_{si}$ are assumed to have an additive structure, the multivariable function $h(z_{1i}, z_{2i}, \dots, z_{si})$ can be written in Eq. (2).

$$h(z_{1i}, z_{2i}, \dots, z_{si}) = h_1(z_{1i}) + h_2(z_{2i}) + \dots + h_s(z_{si}) = \sum_{d=1}^{s} h_d(z_{di}). \tag{2}$$

When each function $h_d(z_{di})$ is approximated by a spline truncated function of degree m with knot points $K_{d1}, K_{d2}, \dots, K_{dr}$, it can be represented in Eq. (3).

$$h_d(z_{di}) = \sum_{j=0}^{m} \vartheta_{dj} z_{di}^{j} + \sum_{k=1}^{r} \vartheta_{d(m+k)} (z_{di} - K_{dk})_{+}^{m} = \vartheta_{d0} + \sum_{j=1}^{m} \vartheta_{dj} z_{di}^{j} + \sum_{k=1}^{r} \vartheta_{d(m+k)} (z_{di} - K_{dk})_{+}^{m} \tag{3}$$

for $i = 1,2, \dots, n$ and $d = 1,2, \dots, s$. The truncated function in Eq. (3) is defined as:

$$(z_{di} - K_{dk})_{+}^{m} = \begin{cases} (z_{di} - K_{dk})^{m} & z_{di} \geq K_{dk} \\ 0 & z_{di} < K_{dk} \end{cases} \tag{4}$$

Substituting Eq. (3) into Eq. (1) yields the multivariable spline truncated nonparametric regression model, expressed in Eq. (5).

$$y_i = \vartheta_0 + \sum_{d=1}^{s} \left(\sum_{j=1}^{m} \vartheta_{dj} z_{di}^{j} + \sum_{k=1}^{r} \vartheta_{d(m+k)} (z_{di} - K_{dk})_{+}^{m} \right) + \varepsilon_i \tag{5}$$

Eq. (5) can be written in matrix form as:

$$\boldsymbol{y} = \boldsymbol{Z}(K)\boldsymbol{\vartheta} + \boldsymbol{\varepsilon} \tag{6}$$

In Eq. (6), $\boldsymbol{y}$ is an $n \times 1$ vector representing the response variabl, $\boldsymbol{Z}(K)$ is an $(n \times (1 + s(m + r))$ design matrix containing the predictor variables included in the model, and $\boldsymbol{\vartheta}$ is a $((1 + s(m + r) \times 1)$ vector of regression coefficients to be estimated. The vector $\boldsymbol{\varepsilon}$ denotes the random error term of dimension $n \times 1$.

2.2. Nonparametric Regression Fourier Series

The Fourier series estimator is a highly flexible trigonometric polynomial, enabling it to optimally adapt to the local characteristics of the data (Ming & Huang, 2018). This method is commonly employed when the underlying data pattern is unknown and exhibits a tendency toward repeated or cyclic behavior. Such repetition refers to the recurring values of the response variable across observations of the predictor variables (Bloomfield, 2000; Prahutama et al., 2018). Suppose that paired data $(v_{1i}, v_{2i}, \dots, v_{qi}, y_i)$ are observed, where the relationship between the predictor variables and the response variable follows a Fourier series nonparametric regression model. The model is expressed in Eq. (7).

$$y_i = g(v_{1i}, v_{2i}, \dots, v_{qi}) + \varepsilon_i \tag{7}$$

The predictor variables $v_{1i}, v_{2i}, \dots, v_{qi}$ are assumed to have an additive structure, such that the multivariable function $g(v_{1i}, v_{2i}, \dots, v_{qi})$ can be written in Eq. (8).

$$g(v_{1i}, v_{2i}, \dots, v_{qi}) = g_1(v_{1i}) + g_2(v_{2i}) + \cdots + g_q(v_{qi}) = \sum_{u=1}^{q} g_u(v_{ui}). \tag{8}$$

When each function $g_u(v_{ui})$ is approximated by a Fourier series with W oscillations, it can be expressed as:

$$g_u(v_{ui}) = \frac{1}{2}\alpha_{0u} + \gamma_u v_{ui} + \sum_{w=1}^{W} \xi_{wu} \cos w\, v_{ui} + \sum_{w=1}^{W} \varphi_{wu} \sin w\, v_{ui} \tag{9}$$

for $i = 1,2,\dots,n$ and $u = 1,2,\dots,q$. Accordingly, the multivariable Fourier series nonparametric regression model can be written as

$$y_i = \frac{1}{2}\alpha_0^* + \sum_{u=1}^{q}\left(\gamma_u v_{ui} + \sum_{w=1}^{W} \xi_{wu} \cos w\, v_{ui} + \sum_{w=1}^{W} \varphi_{wu} \sin w\, v_{ui}\right) + \varepsilon_i \tag{10}$$

where $\alpha_0^* = \sum_{u=1}^{q} \alpha_{0u}$.

Eq. (10) can be compactly expressed in matrix form in Eq. (11).

$$\boldsymbol{y} = \boldsymbol{V}(W)\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{11}$$

In Eq. (11), $\boldsymbol{y}$ is an $(n \times 1)$ vector representing the response variable, $\boldsymbol{V}(W)$ is a design matrix of dimension $(n \times (1 + q + 2qW))$ containing the predictor variables and Fourier oscillation terms used in the analysis, and $\boldsymbol{\alpha}$ is a $((1 + q + 2qW) \times 1)$ vector of regression parameters to be estimated. The vector $\boldsymbol{\varepsilon}$ denotes the random error term of dimension $(n \times 1)$.

2.3. Data Source and Research Variables

This study uses secondary data obtained from the 2023 Indonesian Health Survey (Survei Kesehatan Indonesia, SKI), published by the Ministry of Health of the Republic of Indonesia (Kementerian Kesehatan RI., 2024). The unit of analysis comprises 38 provinces across Indonesia, thereby representing regional-level public health conditions. A purposive sampling technique is employed, with primary considerations given to data availability, recency, and relevance to the research objectives. The selection of the 2023 SKI data is deemed appropriate as it provides comprehensive and up-to-date health indicators, particularly those related to prevalence of heart disease and the associated influencing factors.

Table 1
Research Variables

Variables	Notation	Description	Type of Variables	Number of Observations
Response	y	Prevalence of Heart Disease	Ratio	38
Predictor	x_1	Prevalence of Daily Smoking	Ratio	38
	x_2	Proportion of Hypertension Follow-up/Control at Health Care Facilities	Ratio	38
	x_3	Proportion of Insufficient Physical Activity among Population Aged ≥ 10 Years	Ratio	38
	x_4	Proportion of Habitual Consumption of Fatty/Cholesterol/Fried Foods ≥ 1 Time	Ratio	38

The variables employed in this study consist of a response variable and several predictor variables, which are systematically detailed in Table 1. The selection of predictor variables is based on empirical evidence from the literature as well as theoretical considerations regarding behavioral factors that potentially influence disease prevalence of heart disease. The predictor variables included in this study represent behavioral factors that have been empirically and theoretically shown to contribute significantly to the increased prevalence of heart disease. The relationship between behavioral factors and heart disease prevalence is presumed to be nonlinear in nature; therefore, a flexible nonparametric regression approach is required.

3. Results

The Results section provides a detailed discussion of the development of the Mixed Estimator Nonparametric Regression–Spline Truncated and Fourier Series (MENR–SF) model, beginning with its theoretical formulation and leading to the construction of the model parameter estimation. This section further describes the procedure for selecting the optimal number of spline knot points and Fourier series oscillations using the Generalized Cross-Validation (GCV) criterion developed in this study. In addition, it presents the application of the MENR–SF model to prevalence of heart disease data, including the evaluation of various estimator combinations, until the best-performing model is identified based on estimation accuracy and predictive performance.

3.1 MENR-SF Model Estimation

The initial step is to define the mixed estimator model combining spline truncated and Fourier series (MENR–SF). Suppose paired data are observed with s predictor variables for the spline component and q predictor variables for the Fourier series component, suct that the data can be written as $(z_{1i}, z_{2i}, \dots, z_{si}, v_{1i}, v_{2i}, \dots, v_{qi}, y_i)$ for $i = 1,2,\dots,n$. It is assumed that the relationship between the predictor variables $(z_{1i}, z_{2i}, \dots, z_{si}, v_{1i}, v_{2i}, \dots, v_{qi})$ and the response variables (y_i) follows a mixed nonparametric regression model based on spline truncated and Fourier series estimators (MENR–SF), expressed in Eq. (12).

$$y_i = h(z_{1i}, z_{2i}, \dots, z_{si}) + g(v_{1i}, v_{2i}, \dots, v_{qi}) + \varepsilon_i \tag{12}$$

Based on Eq. (12) and the functional forms of the spline truncated and Fourier series components, Equation (12) can be compactly written as

$$y_i = \mu(z_i, v_i) + \varepsilon_i \tag{13}$$

where:

$$\mu(z_i, v_i) = \sum_{d=1}^{s} h_d(z_{di}) + \sum_{u=1}^{q} g_u(v_{ui})$$

Eq. (13) indicates that the response variable $\boldsymbol{y}$ is influenced by two main components, namely the spline truncated component and the Fourier series component, along with a random error term. This model can be expressed in matrix form as shown in Eq. (14), where $\boldsymbol{Z}(K)\boldsymbol{\vartheta}$ represents the contribution of the spline truncated component and $\boldsymbol{V}(W)\boldsymbol{\alpha}$ represents the contribution of the Fourier series component:

$$\boldsymbol{y} = \boldsymbol{Z}(K)\boldsymbol{\vartheta} + \boldsymbol{V}(W)\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{14}$$

where:

$$\boldsymbol{y} = [y_1 \quad y_2 \quad \cdots \quad y_n]^T$$

$$\boldsymbol{\varepsilon} = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n]^T$$

$$\boldsymbol{Z}(K) = \begin{bmatrix} 1 & z_{11} & \dots & z_{11}^m & (z_{11}-K_{11})_+^m & \dots & (z_{11}-K_{1r})_+^m & \dots & z_{s1} & \dots & z_{s1}^m & (z_{s1}-K_{s1})_+^m & \dots & (z_{s1}-K_{sr})_+^m \\ 1 & z_{12} & \dots & z_{12}^m & (z_{12}-K_{11})_+^m & \dots & (z_{12}-K_{1r})_+^m & \dots & z_{s2} & \dots & z_{s2}^m & (z_{s2}-K_{s1})_+^m & \dots & (z_{s2}-K_{sr})_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{1n} & \dots & z_{1n}^m & (z_{1n}-K_{11})_+^m & \dots & (z_{1n}-K_{1r})_+^m & \dots & z_{sn} & \dots & z_{sn}^m & (z_{sn}-K_{s1})_+^m & \dots & (z_{sn}-K_{sr})_+^m \end{bmatrix}$$

$$\boldsymbol{\vartheta} = \left[\vartheta_0 \quad \vartheta_{11} \quad \cdots \quad \vartheta_{1m} \quad \vartheta_{1(m+1)} \quad \cdots \quad \vartheta_{1(m+r)} \quad \cdots \quad \vartheta_{s1} \quad \cdots \quad \vartheta_{sm} \quad \vartheta_{s(m+1)} \quad \cdots \quad \vartheta_{s(m+r)}\right]^T$$

$$\boldsymbol{V}(W) = \begin{bmatrix} 1/2 & v_{11} & \cos v_{11} & \sin v_{11} & \cdots & \cos Wv_{11} & \sin Wv_{11} & \cdots & v_{q1} & \cos v_{q1} & \sin v_{q1} & \cdots & \cos Wv_{q1} & \sin Wv_{q1} \\ 1/2 & v_{12} & \cos v_{12} & \sin v_{12} & \cdots & \cos Wv_{12} & \sin Wv_{12} & \cdots & v_{q2} & \cos v_{q2} & \sin v_{q2} & \cdots & \cos Wv_{q2} & \sin Wv_{q2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1/2 & v_{1n} & \cos v_{1n} & \sin v_{1n} & \cdots & \cos Wv_{1n} & \sin Wv_{1n} & \cdots & v_{qn} & \cos v_{qn} & \sin v_{qn} & \cdots & \cos Wv_{qn} & \sin Wv_{qn} \end{bmatrix}$$

$$\boldsymbol{\alpha} = \left[\alpha_0^* \quad \gamma_1 \quad \xi_{11} \quad \varphi_{11} \quad \cdots \quad \xi_{W1} \quad \varphi_{W1} \quad \cdots \quad \gamma_q \quad \xi_{1q} \quad \varphi_{1q} \quad \cdots \quad \xi_{Wq} \quad \varphi_{Wq}\right]^T$$

To simplify the model representation and parameter estimation procedure, both components are subsequently combined into a single linear regression form, as given in Equation (15). In this formulation, all information from the spline truncated and Fourier series components is summarized in a single design matrix $\boldsymbol{X}$, while all unknown parameters are collected into a single parameter vector $\boldsymbol{\beta}$.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{15}$$

where:

$$\boldsymbol{X}(K, W) = [\boldsymbol{Z}(K) \quad \boldsymbol{V}(W)]$$
$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\vartheta} \\ \boldsymbol{\alpha} \end{bmatrix}$$

Alternatively, the model parameters can be estimated in a partial manner, in which the parameters associated with the spline truncated component and the Fourier series component are estimated separately. Under this approach, the estimated parameters from each component are subsequently combined through a substitution procedure to obtain the overall mixed estimator. In this study, however, an alternative formulation is adopted by simplifying the mixed estimator into a linear regression structure. This strategy enables the simultaneous estimation of all model parameters using the Least Squares method, thereby improving computational efficiency and ensuring a unified optimization of the estimation process.

In Eq. (15), $\boldsymbol{X}$ is the combined design matrix incorporating all spline truncated and Fourier series components, while $\boldsymbol{\beta}$ is the vector of regression parameters to be estimated. The error vector $\boldsymbol{\varepsilon}$ is assumed to follow a normal distribution with zero mean and constant variance. Parameter estimation for $\boldsymbol{\beta}$ is performed using the Least Squares (LS) method, yielding the estimator in Eq. (16).

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}(K,W)^T\boldsymbol{X}(K,W)\right)^{-1}\boldsymbol{X}(K,W)^T\boldsymbol{y} \tag{16}$$

Under the assumption that $\left(\boldsymbol{X}(K,W)^T\boldsymbol{X}(K,W)\right)^{-1}$ is nonsingular.

This approach allows complex nonparametric models to be efficiently estimated within a linear regression framework, thereby facilitating both implementation and interpretation of the estimation results.

3.2 Construction of Optimal Knot Point and Oscillation Selection

In nonparametric regression using the MENR–SF framework, the selection of optimal spline knot points and the number of Fourier series oscillations can be performed using the Generalized Cross-Validation (GCV) method. GCV is a widely used criterion for determining the optimal number and configuration of knot points and oscillatory components. The GCV formulation developed by Wahba (1990) is presented in Eq. (17).

$$GCV(K,W) = \frac{MSE(K,W)}{\left(n^{-1}trace\left(\boldsymbol{I}-\boldsymbol{H}(K,W)\right)\right)^2} \tag{17}$$

The Mean Squared Error (MSE) is defined as given in Eq. (18).

$$MSE(K,W) = n^{-1}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{18}$$

$\boldsymbol{I}$: identity matrix

$\boldsymbol{H}(K,W)$: $\boldsymbol{X}(K,W)\left(\boldsymbol{X}(K,W)^T\boldsymbol{X}(K,W)\right)^{-1}\boldsymbol{X}(K,W)^T$

The optimal combination of K and W is determined by selecting the values that minimize the GCV criterion, namely:

$$(K^*,W^*) = \underset{K,W}{\arg\min}\ GCV(K,W)$$

This selection yields a MENR–SF model that achieves the best balance between estimation accuracy and model complexity and is therefore expected to provide optimal estimation and predictive performance.

4. Application Example on Real Data

4.1 Descriptive Characteristics of the Data

Table 2 presents the descriptive statistics of the response variable and behavioral predictors related to prevalence of heart disease in Indonesia. The summary statistics provide an overview of the central tendency and variability of each variable, which serves as a preliminary assessment of data characteristics prior to model estimation.

Table 2
Descriptive Statistics

Variable	Minimum	Q1	Median	Mean	Q3	Maximum
y	0.110	0.510	0.640	0.726	0.870	1.670
x_1	10.00	18.23	19.40	19.92	22.68	27.70
x_2	20.40	39.98	42.80	43.78	47.38	59.70
x_3	27.80	38.55	42.85	43.50	50.02	62.00
x_4	12.20	20.32	26.30	27.55	31.43	54.20

The prevalence of heart disease shows substantial variation, with a minimum value of 0.110, a maximum value of 1.670, and a mean of 0.726. This indicates heterogeneity in heart disease prevalence across regions or observation periods. The behavioral predictor variables exhibit varying degrees of dispersion, with the proportion of insufficient physical activity and the habitual consumption of fatty/cholesterol/fried foods displaying relatively wider ranges than the other variables. Furthermore, differences between the median and mean values observed for several variables suggest potential nonlinear relationships between the predictor variables and prevalence of heart disease, thereby supporting the relevance of a nonparametric regression approach in this study.

4.2 Multicollinearity Diagnostics

Multicollinearity assessment is conducted to ensure the absence of strong linear relationships among the predictor variables in the regression model. The indicator used for this purpose is the Variance Inflation Factor (VIF).

Table 3
Variance Inflation Factor (VIF) Results

Predictor Variable	x_1	x_2	x_3	x_4
VIF	1.551	1.011	1.352	1.399

Based on the VIF results, all predictor variables exhibit relatively low VIF values (less than 10), indicating that linear relationships among predictors are minimal and that the model is free from singular or near-singular matrix issues. Therefore, the regression model is considered stable and suitable for subsequent analysis.

5. Exploratory Analysis of Predictor–Response Relationships Using Scatter Plots

The identification of relationship patterns is conducted to obtain an initial understanding of the form of the relationship between the response variable and each predictor variable. Scatter plots are employed to observe relationship trends, data dispersion, and the potential presence of nonlinear patterns prior to further modeling.

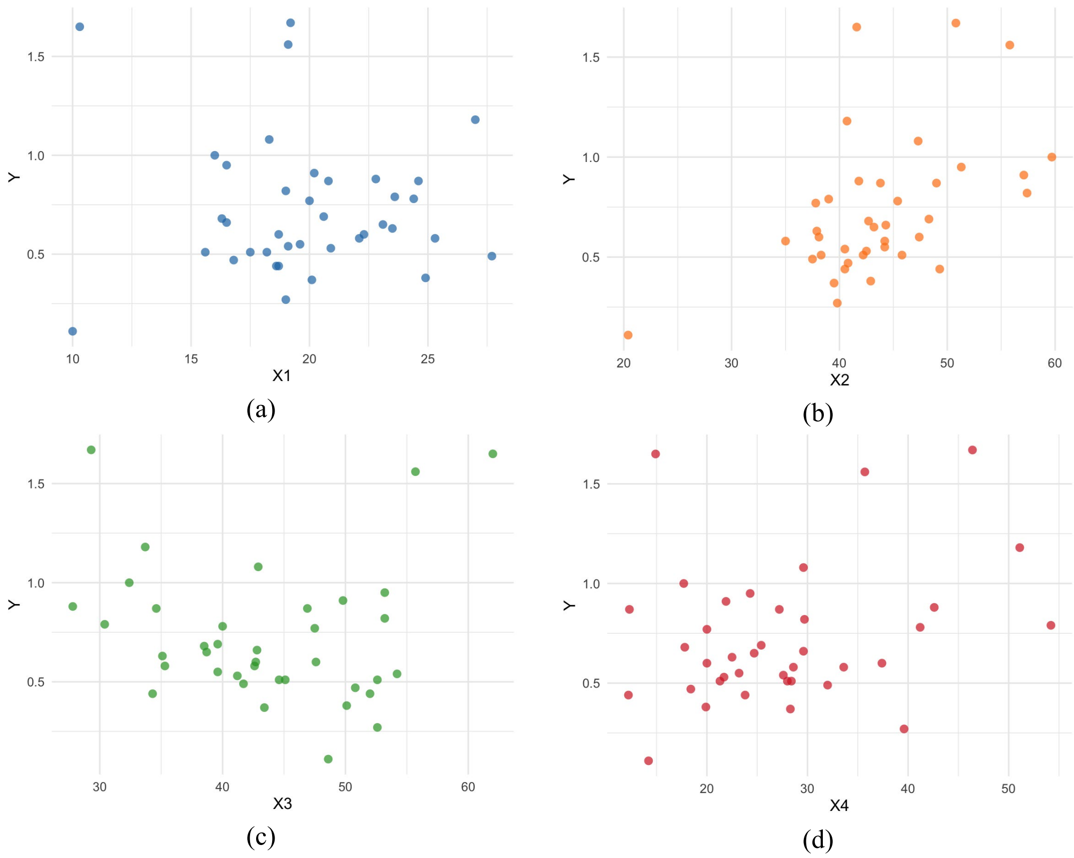

Fig. 1. Relationship patterns between prevalence of heart disease and predictor variables based on scatter plots: (a) x_1 vs. y, (b) x_2 vs. y, (c) x_3 vs. y, and (d) x_4 vs. y

Based on the scatter plots between the response variable (prevalence of heart disease) and each predictor variable, it is observed that each predictor exhibits a distinct relationship pattern. Variable x_1 shows a relationship that varies across several subintervals, suggesting visually that it is more suitably modeled using a spline truncated estimator. Similar patterns are also observed for variables x_3 and x_4, indicating the presence of local structural changes in their relationships with the response variable. In contrast, variable x_2 visually tends to exhibit a fluctuative or oscillatory pattern, which resembles the characteristics of a Fourier series.

Nevertheless, determining the type of estimator solely based on scatter plot visualization may introduce subjectivity and ambiguity. Therefore, this study does not rely exclusively on visual patterns to select estimators. Instead, a comprehensive evaluation of all possible combinations of spline truncated and Fourier series estimators for each predictor variable is conducted. This approach aims to obtain an optimal model structure objectively based on the selected performance criteria. The possible estimator combinations are presented in Table 4.

Table 4
Scheme of Spline truncated and Fourier Series Estimator Combinations for Each Predictor Variable

Combinations	x_1	x_2	x_3	x_4
1	Spline Truncated	Fourier	Fourier	Fourier
2	Fourier	Spline Truncated	Fourier	Fourier
3	Fourier	Fourier	Spline Truncated	Fourier
4	Fourier	Fourier	Fourier	Spline Truncated
5	Spline Truncated	Spline Truncated	Fourier	Fourier
6	Spline Truncated	Fourier	Spline Truncated	Fourier
7	Spline Truncated	Fourier	Fourier	Spline Truncated
8	Fourier	Spline Truncated	Spline Truncated	Fourier
9	Fourier	Spline Truncated	Fourier	Spline Truncated
10	Fourier	Fourier	Spline Truncated	Spline Truncated
11	Spline Truncated	Spline Truncated	Spline Truncated	Fourier
12	Spline Truncated	Spline Truncated	Fourier	Spline Truncated
13	Spline Truncated	Fourier	Spline Truncated	Spline Truncated
14	Fourier	Spline Truncated	Spline Truncated	Spline Truncated

Given that each predictor variable can be modeled using one of two estimators, namely, spline truncated or Fourier series, the total number of possible model configurations for four variables is $2^4 = 16$. However, two extreme configurations, namely the model that employs spline truncated for all predictors and the model that relies entirely on Fourier series for all predictors, are excluded because they do not align with the analytical objective of this study, which emphasizes the proposed mixed-estimator concept. Consequently, a total of 14 model combinations is systematically evaluated using performance measures such as GCV, MSE, RMSE, and R^2. This approach ensures that the selection of the best model is objective, does not rely solely on visual interpretation, and reflects a comprehensive exploration of all methodologically feasible nonparametric model structures.

6. Parameter Estimation of the MENR-SF Model for Prevalence of Heart Disease

Parameter estimation is conducted for each of the 14-spline truncated–Fourier series model combinations evaluated in this study. For the spline truncated component, one to three knot points are examined using a linear order (degree 1), ensuring that the spline remains simple while retaining sufficient flexibility to capture local slope changes. For the Fourier series component, one to three oscillations are considered to represent potential periodic patterns in the data. The optimal combination of knot points and oscillations for each model configuration is selected based on the minimum GCV criterion. This approach ensures that the selection of nonparametric parameters is objective and consistently applied across all 14 evaluated configurations.

The parameter estimation results indicate that the 13th combination provides the best-performing model, in which variables x_1, x_3, and x_4 are modeled using spline truncated estimators, while variable x_2 is modeled using a Fourier series estimator. Based on the evaluation of various knot and oscillation configurations, the optimal setting is achieved with three knot points and three oscillations, yielding a minimum Generalized Cross-Validation (GCV) value of 1.197.

Table 5
Evaluation Results of the MENR-SF Modeling

Combination 13		x_1	x_2	x_3	x_4
		Spline Truncated	**Fourier**	**Spline Truncated**	**Spline Truncated**
Number of Knots	**Number of Oscillations**	**GCV**	**MSE**	**MAPE**	**R^2**
1	1	1.951	0.027	21.78%	0.763
	2	2.014	0.025	20.26%	0.789
	3	2.261	0.024	18.95%	0.798
2	1	1.535	0.017	16.38%	0.851
	2	1.404	0.013	13.54%	0.885
	3	1.485	0.012	13.96%	0.898
3	1	1.422	0.013	14.96%	0.893
	2	1.244	0.009	11.26%	0.923
	3	**1.197**	**0.007**	**10.65%**	**0.940**

Based on the evaluation results, the best model combination is obtained using three knot points and three oscillations, as indicated by the minimum GCV value of 1.197. The low GCV value suggests that the model achieves an optimal balance between goodness of fit and model complexity, thereby minimizing the risk of overfitting. In addition, this combination yields the lowest MSE and MAPE values, at 0.007 and 10.65%, respectively, indicating relatively low prediction errors. The high coefficient of determination (R^2=0.940) indicates that approximately 94% of the variation in the response variable is explained by the model. Therefore, it can be concluded that the MENR–SF model with three knot points and three oscillations is the most accurate and efficient among the evaluated combinations. The best MENR–SF model is presented in Equation (19).

$$\begin{aligned}\hat{y} = -0.111 - 0.007x_1 + 0.004x_3 - 0.010x_4 - 0.042(x_1 - 20.11)_+ + 0.107(x_1 - 22.64)_+ \\ - 0.114(x_1 - 25.17)_+ - 0.061(x_3 - 47.34)_+ + 0.321(x_3 - 52.23)_+ \\ - 0.347(x_3 - 57.11)_+ + 0.067(x_4 - 36.20)_+ + 0.108(x_4 - 42.20)_+ \\ - 0.297(x_4 - 48.20)_+ - 0.055 + 0.0232x_2 + 0.062\cos x_2 - 0.116\sin x_2 \\ + 0.078\cos 2x_2 + 0.111\sin 2x_2 + 0.020\cos 3x_2 - 0.125\sin 3x_2\end{aligned} \quad (19)$$

Based on the best MENR–SF model presented in Eq. (19), the interpretation of the results is summarized in Table 6.

Table 6
Interpretation of predictor effects in the best MENR–SF model

Variable	Best Estimator	Interpretation
x_1 (Prevalence of Daily Smoking)	Spline truncated	The initial linear coefficient of x_1 is negative (−0.007), indicating that at smoking prevalence levels below 20.11%, an increase in smoking prevalence does not significantly increase prevalence of heart disease. However, after exceeding the second knot (22.64%), the spline coefficient becomes strongly positive (+0.107), suggesting that increases in smoking prevalence within the intermediate range substantially raise prevalence of heart disease. Beyond 25.17%, the coefficient becomes negative again (−0.114), indicating a saturation effect, where further increases in smoking prevalence no longer proportionally increase the risk.
x_2 (Proportion of Hypertension Follow-up/Control at Health Care Facilities)	Fourier Series	The linear coefficient of x_2 is small and positive (+0.0232), while the sine and cosine Fourier components exhibit alternating signs. This pattern indicates that the relationship between hypertension control coverage and prevalence of heart disease is non-monotonic and exhibits a fluctuating behavior.
x_3 (Proportion of Insufficient Physical Activity among Population Aged ≥ 10 Years)	Spline truncated	The initial linear coefficient of x_3 is positive (+0.004), indicating that at relatively low levels of insufficient physical activity, an increase in the proportion of physically inactive individuals leads to a higher prevalence of heart disease. The spline coefficient at the second knot (52.23%) is strongly positive (+0.321), suggesting a sharp increase in heart disease prevalence when more than half of the population has low physical activity. Conversely, the negative coefficient at the third knot (−0.347) indicates a slowdown in the marginal effect at very high levels of physical inactivity.
x_4 (Proportion of Habitual Consumption of Fatty/Cholesterol/Fried Foods ≥ 1 Time)	Spline truncated	The initial linear coefficient of x_4 is negative (−0.010), whereas the spline coefficients at the first two knots are positive (+0.067 and +0.108). This indicates that at moderate consumption levels, increased intake of fatty foods significantly raises prevalence of heart disease. After exceeding 48.20%, the coefficient becomes strongly negative (−0.297), reflecting a saturation effect in which further increases in consumption no longer substantially elevate prevalence.

After completing parameter estimation and interpretation, the next step is to evaluate the predictive performance of the model. Fig. 2 illustrates the agreement between the predicted values and the observed values obtained from the best MENR–SF model.

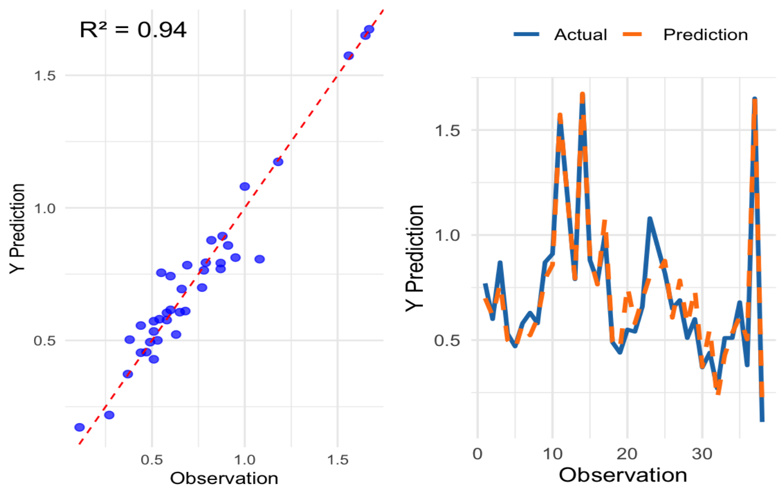

Fig. 2. Comparison between observed and predicted values of the best MENR–SF model

Fig. 2 presents a comparison between the observed values and the predicted values generated by the best-performing model. The left panel shows a scatter plot of observed versus predicted values, with the diagonal line serving as a reference for perfect agreement. The concentration of points near the diagonal line, together with the high coefficient of determination ($R^2 = 0.94$), indicates that the model has a very strong ability to explain data variability and to produce accurate predictions. The right panel displays a comparison of the patterns between the observed and predicted values across the observation sequence. The predicted curve closely follows the fluctuations of the observed data, including sharp changes in values. This demonstrates that the model is not only accurate in an aggregate sense but is also effective in capturing local dynamics and structural variations within the data. Overall, these results confirm that the best MENR–SF model provides optimal modeling performance in terms of both predictive accuracy and pattern conformity.

6. Discussion

This study demonstrates that the Mixed Estimator Nonparametric Regression–Spline Truncated–Fourier Series (MENR–SF) constitutes a flexible modeling framework for capturing complex nonlinear relationships between behavioral factors and prevalence of heart disease in Indonesia. The discussion focuses on both the theoretical development of the model and its implementation in a real-world application.

The empirical results indicate that MENR–SF effectively accommodates data characteristics by allowing each predictor variable to be modeled using the most appropriate estimator. The selection of the optimal model configuration is conducted objectively using the Generalized Cross-Validation (GCV) criterion, thereby achieving an optimal balance between estimation accuracy and model complexity. The best-performing configuration—Combination 13, comprising three spline truncated predictors and one Fourier series predictor, yields a minimum GCV value of 1.197, low prediction errors (MSE = 0.007; MAPE = 10.65%), and a high coefficient of determination ($R^2 = 0.94$). These results demonstrate the strong capability of MENR–SF to explain variations in prevalence of heart disease data. Moreover, the use of the Least Squares (LS) approach for parameter estimation enables the complex MENR–SF model to be estimated efficiently and stably. Integrating the spline truncated and Fourier series components into a single linear regression design matrix further facilitates implementation and enhances the model's practical applicability to regional-scale health data.

This approach extends previous studies that typically applied fixed estimator combinations. The findings emphasize that selecting estimators based solely on scatter plot visualization may be subjective; therefore, a systematic evaluation of all possible estimator combinations using performance criteria is essential. By evaluating 14 feasible model combinations, MENR–SF ensures that the final model structure genuinely reflects the underlying data characteristics. This strategy strengthens the position of MENR–SF as a significant methodological advancement in multivariable nonparametric regression.

A key advantage of MENR–SF lies in its ability to reveal heterogeneous relationship patterns between individual behavioral factors and prevalence of heart disease. The prevalence of daily smoking (x_1), modeled using spline truncated, exhibits a threshold effect increases in smoking prevalence at intermediate levels have a stronger impact on prevalence of heart disease, whereas at very high levels, further increases do not proportionally elevate risk. Similar patterns are observed for the proportion of insufficient physical activity (x_3) and the habitual consumption of fatty/cholesterol/fried foods (x_4). Large spline coefficients beyond certain knot points indicate sharp risk increases once critical behavioral thresholds are exceeded. Conversely, negative coefficients at very high levels suggest a deceleration of marginal effects, potentially reflecting behavioral homogeneity or limited data variability at extreme levels.

In contrast to the other variables, the proportion of hypertension follow-up or control at health care facilities (x_2) is most appropriately modeled using a Fourier series estimator. The resulting oscillatory pattern indicates a non-monotonic and fluctuating relationship between hypertension control coverage and prevalence of heart disease. This behavior may reflect regional differences in health service effectiveness, delayed impacts of hypertension management, or other systemic factors with periodic characteristics that cannot be adequately captured by spline-based estimators alone.

7. Conclusion

This study successfully introduces a methodological innovation through the application of the Mixed Estimator Nonparametric Regression–Spline Truncated Fourier Series (MENR–SF) for modeling nonlinear relationships in prevalence of heart disease. This approach allows the simultaneous combination of spline truncated and Fourier series estimators, thereby providing greater flexibility in estimation. The model's accuracy is evidenced by relatively low error values and a high coefficient of determination, indicating a strong ability to explain data variability. The modeling process is conducted through the selection of the optimal number of spline knot points and Fourier oscillations based on the Generalized Cross-Validation (GCV) criterion, ensuring that the resulting model not only achieves a good fit to the data but also remains efficient in terms of model complexity.

8. Limitations and Future Research Directions

This study has several limitations that should be acknowledged. The analysis is primarily limited to the parameter estimation process; therefore, the results focus on the model's ability to fit the data without incorporating formal statistical inference. As a direction for future research, the development of the MENR–SF model may be extended to include inferential aspects, particularly through simultaneous and partial hypothesis testing of the model parameters. Such an extension is expected to provide stronger insights into the statistical significance of the effects of individual variables, thereby enabling the model to be not only robust in terms of estimation and predictive accuracy but also supported by a more comprehensive inferential framework for data-driven decision making.

Author Contributions

Conceptualization, I.N.B., N.C., and A.T.R.D; methodology, I.N.B..; software, I.N.B., and A.T.R.D; validation, I.N.B. and N.C.; formal analysis, I.N.B. and A.T.R.D; investigation, I.N.B; resources, N.C. and M.A.; data curation, A.T.R.D; writing—original draft preparation, I.N.B., N.C., and A.T.R.D; writing—review and editing, N.C. and M.A; visualization, A.T.R.D; supervision, I.N.B. and N.C; project administration, I.N.B. and N.C; funding acquisition, I.N.B. and N.C. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgement

This research is funded by the Indonesian Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science and Technology and managed under the EQUITY Program (Contract No. 4299/B3/DT.03.08/2025 & No 3029/PKS/ITS/2025.

References

Amri, I. F., Chamidah, N., Saifudin, T., Purwanto, D., Fadlurohman, A., Fitriyana Ningrum, A., & Amri, S. (2024). Prediction of extreme weather using nonparametric regression approach with Fourier series estimators. *Data and Metadata*, *4*, 1–12. https://doi.org/10.56294/dm2024319

Bilodeau, M. (1992). Fourier smoother and additive models. *The Canadian Journal of Statistics*, *20*(3), 257–269.

Bloomfield, P. (2000). *Fourier Analysis of Time Series An Introduction* (2nd ed.). John Wiley & Sons, Inc.

Budiantara, I. N. (2019). *Regresi Nonparametrik Spline Truncated* (1st ed.). ITS Press.

Budiantara, I. N., Ratnasari, V., Ratna, M., & Zain, I. (2015). The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties. *Applied Mathematical Sciences*, *9*(122), 6083–6094. https://doi.org/10.12988/ams.2015.58517

Chamidah, N., Lestari, B., Susilo, H., Dewi, T. K., Saifudin, T., Al Akhwal Siregar, N. R., & Aydin, D. (2025). Modeling coronary heart disease risk based on age, fatty food consumption and anxiety factors using penalized spline nonparametric logistic regression. *MethodsX*, *14*. https://doi.org/10.1016/j.mex.2025.103320

Chen, H. (1991). Polynomial splines and nonparametric regression. *Journal of Nonparametric Statistics*, *1*(1–2), 143–156. https://doi.org/10.1080/10485259108832516

Chu, C.-K., & Marron, J. S. (1991). Choosing a Kernel Regression Estimator. *Statistical Science*, *6*(4), 404–436.

Cui, W., & Wei, M. (2013). Strong Consistency of Kernel Regression Estimate. *Open Journal of Statistics*, *03*(03), 179–182. https://doi.org/10.4236/ojs.2013.33020

Dani, A. T. R., Ratnasari, V., & Budiantara, I. N. (2021). Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression. *IOP Conference Series: Materials Science and Engineering*, *1115*(1), 012020. https://doi.org/10.1088/1757-899x/1115/1/012020

Eilers, P. H. C., & Marx, B. D. (2010). Splines, Knots, and Penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(6), 637–653. https://doi.org/10.1002/wics.125

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. Marcel Dekker.

Fatmawati, F., Nyoman Budiantara, I., & Lestari, B. (2019). Comparison of Smoothing and Truncated Spline Estimators in Estimating Blood Pressure Models. *International Journal of Innovation, Creativity, and Change*, *5*(3), 1177–1199. www.ijicc.net

Ganesh, A., Balasubramanian, G., Jena, S. K., & Pradhan, N. (2011). Fourier Approach to Function Approximation. *International Journal of Mathematical Archive*, *4*, 617–624.

Greblicki, W., & Pawlak, M. (1985). Fourier and Hermite series estimates of regression functions. *Annals of the Institute of Statistical Mathematics*, *37*(1), 443–454. https://doi.org/10.1007/BF02481112

Hackshaw, A., Morris, J. K., Boniface, S., Tang, J. L., & Milenkovi, D. (2018). Low cigarette consumption and risk of coronary heart disease and stroke: Meta-analysis of 141 cohort studies in 55 study reports. In *BMJ (Online)* (Vol. 360). BMJ Publishing Group. https://doi.org/10.1136/bmj.j5855

Hajduk, A. M., & Chaudhry, S. I. (2016). Sedentary Behavior and Cardiovascular Risk in Older Adults: a Scoping Review. In *Current Cardiovascular Risk Reports* (Vol. 10, Issue 1, pp. 1–11). Current Medicine Group LLC 1. https://doi.org/10.1007/s12170-016-0485-6

Hardle, W. H., & Kelly, G. K. (1987). Nonparametric Kernel Regression Estimation-Optimal Choice of Bandwidth. *Statistics*, *18*(1), 21–35. https://doi.org/10.1080/02331888708801986

Huang, M., Li, R., & Wang, S. (2013). Nonparametric mixture of regression models. *Journal of the American Statistical Association*, *108*(503), 929–941. https://doi.org/10.1080/01621459.2013.772897

Kementerian Kesehatan RI. (2024). *SURVEI KESEHATAN INDONESIA (SKI) 2023 DALAM ANGKA*. Kementerian Kesehatan Republik Indonesia.

Kementerian Kesehatan RI. (2025). *Profil Kesehatan Indonesia 2024*. Kementerian Kesehatan Republik Indonesia.

Kim, A. S., & Johnston, S. C. (2011). Global variation in the relative burden of stroke and ischemic heart disease. *Circulation*, *124*(3), 314–323. https://doi.org/10.1161/CIRCULATIONAHA.111.018820

Lestari, B., Fatmawati, F., & Budiantara, I. N. (2020). Spline Estimator and Its Asymptotic Properties in Multiresponse Nonparametric Regression Model. *Songklanakarin Journal of Science and Technology*, *42*(3), 533–548.

Liu, H., Liu, Z., Gong, Y., Guo, J., Liu, X., Sun, Y., Tang, W., Cheng, W., & Jin, W. (2025). Low physical activity-related disease burden, 1990-2021: assessment of global trends and social determinants based on GBD 2021 data. *Journal of Global Health*, *15*, 04314. https://doi.org/10.7189/jogh.15.04314

Ming, W. Y., & Huang, L.-J. (2018). Fourier Series Neural Networks for Regression. *Proceedings of IEEE International Conference on Applied System Innovation*, 716–719.

Prahutama, A., Suparti, & Utami, T. W. (2018). Modelling fourier regression for time series data - A case study: Modelling inflation in foods sector in Indonesia. *Journal of Physics: Conference Series*, *974*(1). https://doi.org/10.1088/1742-6596/974/1/012067

Ramli, M., Ratnasari, V., & Nyoman Budiantara, I. (2020). Estimation of Matrix Variance-Covariance on Nonparametric Regression Spline Truncated for Longitudinal Data. *Journal of Physics: Conference Series*, *1562*(1). https://doi.org/10.1088/1742-6596/1562/1/012014

Ratnasari, V., Budiantara, I. N., Ratna, M., & Zain, I. (2016). Estimation of Nonparametric Regression Curve using Mixed Estimator of Multivariable Truncated Spline and Multivariable Kernel. *Global Journal of Pure and Applied Mathematics*, *12*(6), 5047–5057. http://www.ripublication.com/gjpam.htm

Ratnasari, V., Budiantara, N., & Dani, A. T. R. (2021). Nonparametric Regression Mixed Estimators of Truncated Spline and Gaussian Kernel based on Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR) Methods. *International Journal on Advanced Science Engineering Information Technology*, *11*(6), 2400–2406.

Sanchis-Gomar, F., Perez-Quilis, C., Leischik, R., & Lucia, A. (2016). Epidemiology of coronary heart disease and acute coronary syndrome. In *Annals of Translational Medicine* (Vol. 4, Issue 13). AME Publishing Company. https://doi.org/10.21037/atm.2016.06.33

Wahba. (1990). *Spline Models for Observational Data* (2nd ed.). SIAM.

Wahba, G., & Wang, Y. (2014). Spline Function. *Encyclopedia of Statistical Sciences*, 1–27. https://doi.org/10.4135/9781446247501.n3679

Yatchew, A. (1998). Nonparametric Regression Techniques in Economics. *Journal of Economic Literature*, *36*(2), 669–721. http://www.jstor.orgURL:http://www.jstor.org/stable/2565120http://www.jstor.org/page/info/about/policies/terms.jsp

Zhang, X., Ru, J., & Wu, C. (2023). A Nonparametric Regression-Based Multi-Scale Gradient Correlation Filtering Method for Infrared Small Target Detection. *Electronics (Switzerland)*, *12*(7). https://doi.org/10.3390/electronics12071562